ANNALS OF SURGERY
Vol. 222, No. 6, 735–742

The Objective Structured Clinical Examination

The New Gold Standard for Evaluating Postgraduate Clinical Performance

David A. Sloan, M.D., Michael B. Donnelly, Ph.D., Richard W. Schwartz, M.D., and William E. Strodel, M.D.

From the Department of Surgery, University of Kentucky Chandler Medical Center, Lexington, Kentucky

Objective
The authors determine the reliability, validity, and usefulness of the Objective Structured Clinical Examination (OSCE) in the evaluation of surgical residents.

Summary Background Data
Interest is increasing in using the OSCE as a measurement of clinical competence and as a certification tool. However, concerns exist about the reliability, feasibility, and cost of the OSCE. Experience with the OSCE in postgraduate training programs is limited.

Methods
A comprehensive 38-station OSCE was administered to 56 surgical residents. Residents were grouped into three levels of training: interns, junior residents, and senior residents. The reliability of the examination was assessed by coefficient α; its validity, by the construct of experience. Differences between training levels and in performance on the various OSCE problems were determined by a three-way analysis of variance with two repeated measures and the Student-Newman-Keuls post hoc test. Pearson correlations were used to determine the relationship between OSCE and American Board of Surgery In-Training Examination (ABSITE) scores.

Results
The reliability of the OSCE was very high (0.91). Performance varied significantly according to level of training (postgraduate year; $p < 0.0001$). Senior residents performed best, and interns performed worst. The OSCE problems differed significantly in difficulty ($p < 0.0001$). Overall scores were poor. Important and specific performance deficits were identified at all levels of training. The ABSITE clinical scores, unlike the basic science scores, correlated modestly with the OSCE scores when level of training was held constant.

Conclusion
The OSCE is a highly reliable and valid clinical examination that provides unique information about the performance of individual residents and the quality of postgraduate training programs.

The primary goal of training programs is to produce competent practitioners. In most training programs, the performance of residents is judged by subjective faculty evaluations and standardized multiple choice tests, such as the In-Training Examination. This type of assessment program is problematic for two reasons: (1) Subjective faculty evaluations are unreliable[1-3] and tend to inflate resident performance[3-5]; and (2) multiple-choice examinations, such as the In-Training Examination or the appropriate American Board Certifying Examination, although very reliable, assess only a single dimension of clinical competence, that is, knowledge base.[6] Other important aspects of clinical expertise, such as physical examination skills, interpersonal skills, technical skills, problem-solving abilities, decision-making abilities, and patient treatment skills are not assessed objectively.

Recently, clinicians have focused on the Objective Structured Clinical Examination (OSCE), a multidimensional practical examination of clinical skills, as a tool for assessing clinical competence.[7-11] Although most of the information on the OSCE has been gained from experience with medical students, a handful of studies have shown the value of this tool in assessing the performance of residents.[9-12] Experience with the OSCE is somewhat limited in the United States, but this evaluative method has emerged elsewhere as the premiere method for assessing clinical competence. In Canada, for example, all physicians must now pass the OSCE if they are to be licensed by the Medical Council of Canada.[13,14] Legitimate concerns linger, however, about the wide range of reliabilities (0.19–0.89) reported for the OSCE[9,10,14-18]; there is no universal agreement that conventional assessment methods must be abandoned and replaced by such performance-based tests.

A pilot study at our institution demonstrated that the OSCE reliably measured the clinical performance of interns.[11] We have also shown that the OSCE furnishes information on clinical competence distinct from that provided by faculty ward evaluations and the American Board of Surgery In-Training Examination (ABSITE).[19] Numerous competency deficits are uncovered by the OSCE, and we and others believe that medical graduates enter postgraduate training programs with weak, and even declining, clinical skills.[11,20-23]

In the current study we had three goals: (1) To determine the *reliability* of the OSCE in testing an entire residency population; (2) to determine the *validity* of the OSCE in measuring the performance of residents at multiple training levels; and (3) to determine the usefulness of the information gained about residents' clinical skills. To our knowledge, such an assessment of an entire cohort of residents has never been performed.

METHODS

A comprehensive 38-station OSCE was administered to 56 surgical residents whose postgraduate year (PGY) levels ranged from PGY1 to PGY6. The examination was administered just before the beginning of the academic year, so that the PGY1 residents (interns) were entering the training program. All residents were from the same surgical residency program at the University of Kentucky. For the purposes of data analysis, the residents were grouped into three levels of training: (1) incoming interns (n = 18); (2) junior residents (PGY2 and PGY3; n = 25); and (3) senior residents (PGY4, PGY5, and PGY6; n = 13). The examination was conducted on two consecutive Saturday mornings; half of the residents were evaluated on each day.

The OSCE consisted of 19 clinical problems (Table 1); each clinical problem was divided into two parts (A and B). Part A consisted of a 5-minute interaction between patient and resident, in which the resident was usually asked either to obtain a focused history or to perform a physical examination. Occasionally, the resident was asked to perform a technical exercise (i.e., suturing a laceration) or to give a second opinion (i.e., explain the options to a patient with newly diagnosed breast cancer). Some of the patients used were actual patients, whereas others were playing the role of a patient. A faculty member graded each resident according to a given set of predetermined criteria presented in the form of a checklist. The items on the checklist were the key history or physical examination items that had been deemed by expert faculty to be critical to a competent performance. For example, a breast surgeon listed all of the key physical examination maneuvers and the pertinent clinical findings that he considered essential to making a diagnosis in the case of a patient with fibrocystic breast disease. At each examination station, the faculty members acted as passive evaluators and were instructed not to guide or prompt the residents. Although surgical residents were being evaluated, several faculty members were recruited from other departments, including internal medicine, to proctor those stations that dealt with clinical problems of a broader nature. A total of 34 faculty members participated in the examination.

Part B of each clinical problem consisted of a series of questions dealing directly with the patient interaction completed in part A. For example, having just examined a patient with a thyroid nodule, the residents were asked to respond to questions concerning the diagnostic

Supported in part by a grant from the Association for Surgical Education.

Address reprint requests to David A. Sloan, M.D., Department of Surgery, University of Kentucky Chandler Medical Center, 800 Rose Street, Lexington, KY 40536-0084.

Accepted for publication November 23, 1994.

Table 1. CONTENT OF THE RESIDENT OBJECTIVE STRUCTURED CLINICAL EXAMINATION (OSCE)

OSCE Problem	Description
Arterial ischemia	Examination of an actual patient with signs of advanced peripheral vascular disease
Knee trauma	Examination of a simulated patient with a history of knee trauma
Biliary colic	Evaluation of a simulated patient with symptomatic cholelithiasis; explanation of options
Suturing	Suturing a standardized laceration on a pig's hock
Sciatica	Examination of a simulated patient with back pain
Postoperative abdominal pain	Evaluation of a simulated postoperative patient with new onset of abdominal pain and fever
Breast cancer options	Explanation of treatment options to a patient with newly diagnosed breast cancer
Multitrauma assessment	Resuscitation of a simulated patient with multiple injuries
Tongue cancer	Obtaining a history from a simulated patient with a new sore on his tongue
Hypercalcemia	Assessment of an actual patient with symptoms of primary hyperparathyroidism
Thyroid nodule	Examination of an actual patient with a solitary thyroid nodule
Venous leg ulcer	Examination of an actual patient with a venous stasis ulcer
Surgical anatomy	Identification of multiple surface anatomy landmarks of surgical importance
Hematuria	Evaluation of a simulated patient with gross hematuria
Breast examination	History and physical examination of an actual patient with fibrocystic breast disease
ICU	Paper case dealing with a complex ICU patient management problem
Abdominal pain	Obtaining a history from a patient with left lower quadrant pain and rectal bleeding
Mole evaluation	Focused history and physical examination of an actual patient with a mole
Lung cancer	Obtaining a history from a simulated patient with symptomatology suggestive of lung cancer

ICU = intensive care unit.

workup and treatment of the patient. The part B answers were graded according to a checklist of objective criteria again preset by expert faculty members. The part B stations were also 5 minutes in duration.

A computerized buzzer system signaled the residents to rotate from station to station until each candidate had visited every station. The total examination time for each resident was 190 minutes.

The reliability of the examination was assessed by coefficient alpha. Reliability coefficients were calculated for part A, part B, and combined (averaged) part A/part B. The unit of analysis in calculating the reliability coefficients was the total score. The number of stations needed for the benchmark reliability (0.80) was estimated by the Spearman–Brown formula. Validity was assessed with use of the construct of experience (e.g., a senior resident should perform better than junior residents). A three-way analysis of variance with two repeated measures and the Student-Newman-Keuls *post hoc* test were used to determine whether there were significant differences among PGY levels, whether there was a significant difference in performance on part A and part B, and whether there were differences in performance among the various clinical problems. Pearson's correlations were used to determine the magnitude of the relationship between level of performance and level of training and to determine the relationship between the OSCE and the ABSITE. Competent performance on the OSCE was operationally defined as 60%. This standard applied to individual problems as well as to the overall examination. Considering that this was the first OSCE that the entire group of residents had taken, the decision to set 60% as a passing score for each station was arbitrary. The 0.05 level of confidence was used to define a significant difference.

RESULTS

The reliability of the part A component of the examination was 0.87. The reliability of part B was 0.83. When parts A and B were combined, the reliability of the entire examination increased to 0.91. The Spearman–Brown formula indicated that an 8-problem, 16-station OSCE was required to reach the benchmark reliability of 0.80.

Performance varied significantly by level of training ($F = 53.87$; $df = 2, 53$; $p < 0.0001$). The Student-Newman-Keuls *post hoc* test indicated that the three resident groups were significantly different from each other: the senior residents performed best; the interns, worst. Pearson's correlation between level of training and average OSCE percentage score was 0.80. All of these data strongly support the construct validity of the OSCE.

The residents performed better on part A than on part B ($F = 220.18$; $df = 1, 53$; $p < 0.0001$). Figure 1 shows the mean OSCE part A and part B percentage scores for each resident group. This figure shows that there is an almost constant difference between part A and part B for each of the three groups. There was also a significant

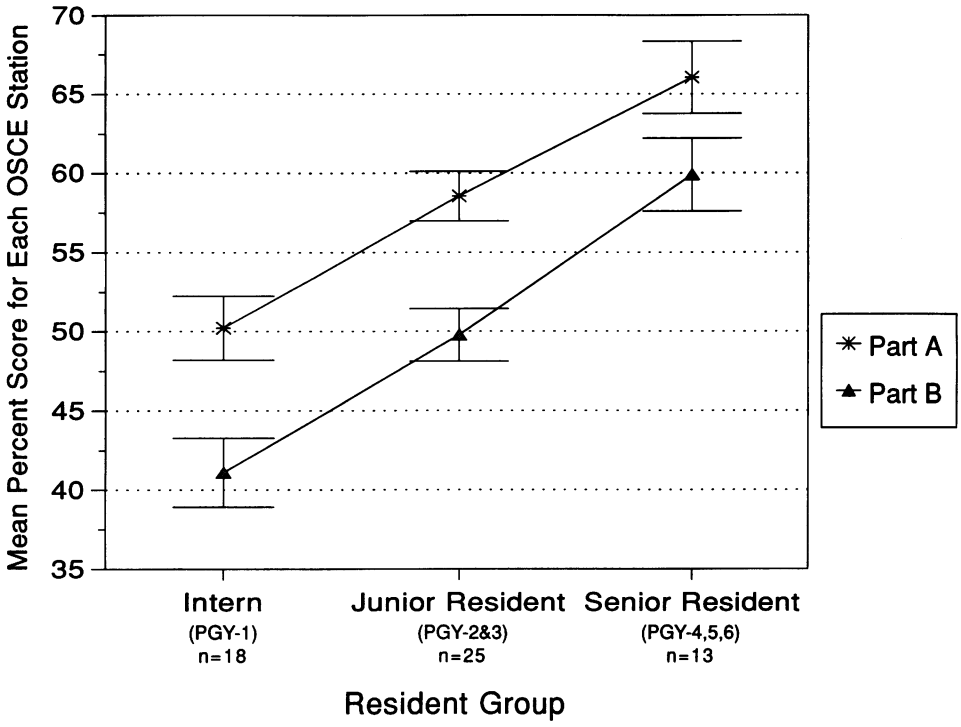

Figure 1. Mean percentage OSCE score and 95% confidence interval for each group.

difference in the difficulty of the various OSCE problems ($F = 39.89$; $df = 18, 954$; $p < 0.0001$). Figure 2 shows the mean percentage score for each OSCE problem. As can be seen from this figure, the means varied from 37% for the arterial ischemia problem to 73% for the lung cancer problem.

Forty-three of the 56 residents also took the ABSITE examination the following January. Performance on the OSCE was correlated with performance on the ABSITE. The results of these analyses are shown in Table 2. The second column of the table presents the simple correlations between the OSCE and ABSITE scores. The ABSITE Clinical score correlated very highly with the three OSCE scores. In contrast, the ABSITE Basic Science score correlated much more modestly with the OSCE scores. We hypothesized that these correlations might be artificially high because scores on both the OSCE and the ABSITE systematically vary by level of training. To control for this factor, we computed partial correlations between the OSCE and ABSITE scores, holding level of training constant. These correlations are presented in the third column of Table 2. As can be seen from the table, the partial correlations are dramatically lower than the simple correlations. The ABSITE Clinical score correlates most highly with the OSCE scores, although these correlations are not very high.

Figure 3 shows the distribution of the average OSCE problem scores. The scores ranged from the 30s to the 60s. Seventy-five percent of the scores were below the established acceptable performance level of 60%. As noted earlier, the 60% passing standard was arbitrary and perhaps not reflective of competence. Items on the checklist varied in importance. In conducting future OSCEs, we plan to prospectively determine key competence items and then base a passing score on these key items. The percentage of residents whose performance was deficient on each OSCE problem is shown in Figure 4. As can be seen from this figure, significant group deficits existed for most of the OSCE problems.

The direct cost for the one-time examination was calculated to be $74.00 per resident. The facility and faculty time were donated. Patients were paid $20.00 per hour for their training and participation in the OSCE.

DISCUSSION

Physicians recognize that clinical competence is determined by more than knowledge. Although a sound knowledge base is vital, clinical competence encompasses numerous other domains, including interviewing and interpersonal skills, physical diagnosis skills, problem-solving abilities, and technical skills. Unfortunately, many of the skills crucial to the competent performance of a physician or surgeon are poorly evaluated by faculty members.[3,24] Stillman et al. noted that in many cases, internal medicine residents taking a history or performing a physical examination were never observed by faculty members.[24] Researchers have established that the reliability and validity of faculty rating forms are generally poor and that these ratings do not correlate well with

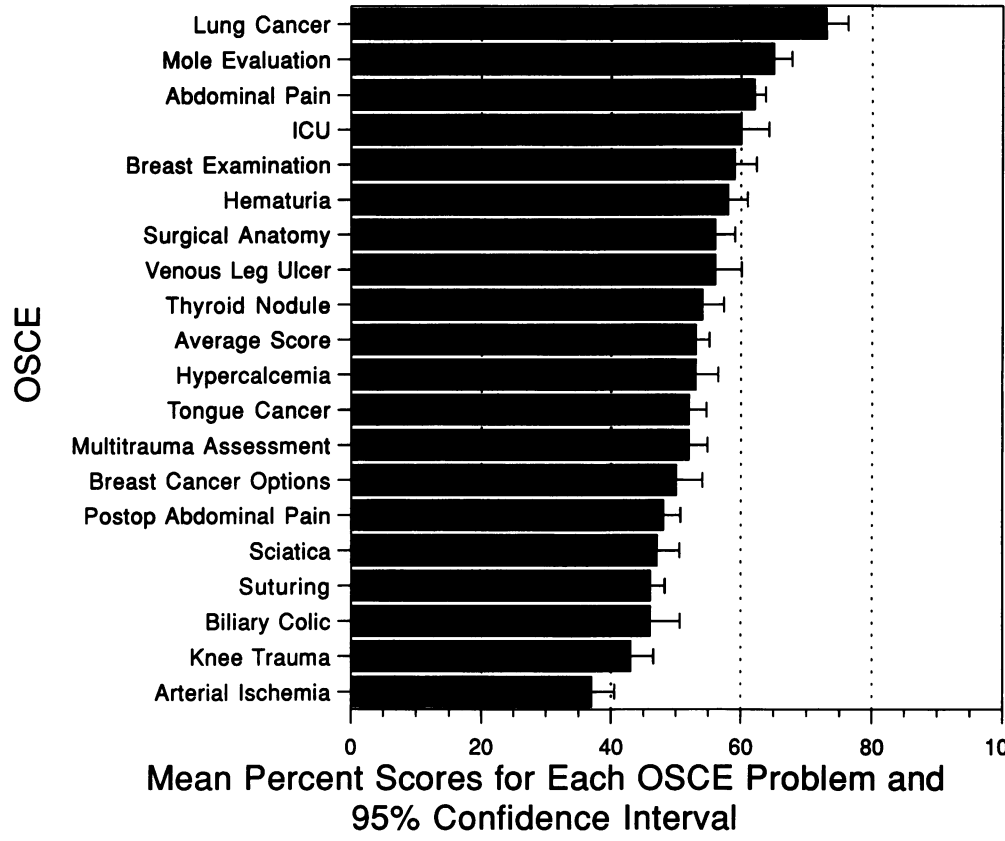

Figure 2. Mean percentage score for each OSCE station.

more objective measures of clinical competence.[1-3,16,19] Other studies have noted that faculty members typically inflate resident performance and are generally reluctant to underscore deficits in clinical performance.[4,5,19] Multiple-choice tests, such as the ABSITE and the tests developed by the National Board of Medical Examiners, although very reliable for measuring knowledge, do not

Table 2. CORRELATION OF OBJECTIVE STRUCTURED CLINICAL EXAMINATION SCORES WITH AMERICAN BOARD OF SURGERY IN-TRAINING EXAMINATION PERFORMANCE

	Correlation	Partial Correlation
ABSITE total		
OSCE total	0.66*	0.25
OSCE part A	0.58*	0.09
OSCE part B	0.71*	0.37†
ABSITE clinical		
OSCE total	0.81*	0.42†
OSCE part A	0.78*	0.37†
OSCE part B	0.80*	0.43‡
ABSITE basic science		
OSCE total	0.39†	0.10
OSCE part A	0.28	−0.08
OSCE part B	0.48‡	0.27

ABSITE = American Board of Surgery In-Training Examination; OSCE = Objective Structured Clinical Examination.
* $p < 0.0001$.
† $p < 0.05$.
‡ $p < 0.01$.

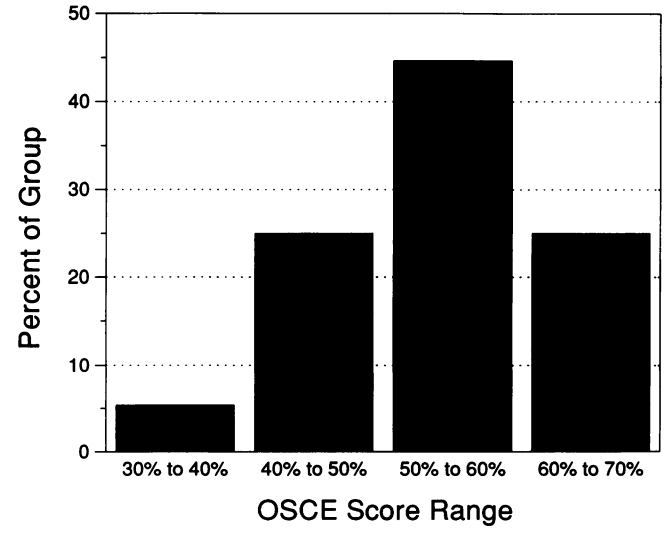

Figure 3. Distribution of average OSCE scores.

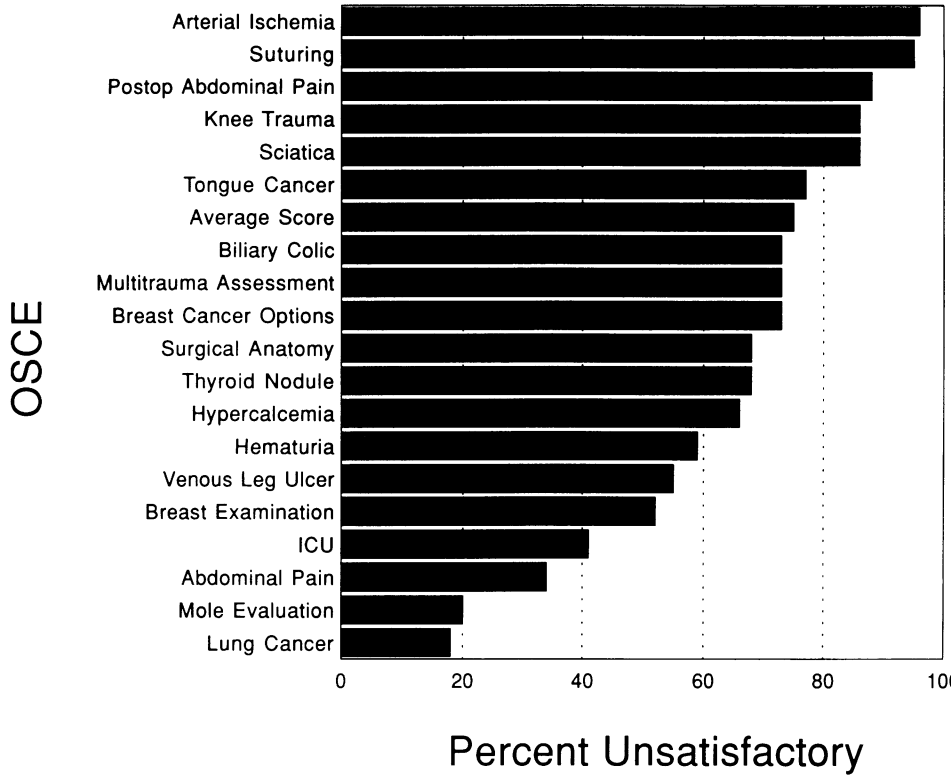

Figure 4. Percentage of residents scoring less than 60%.

measure clinical ability, nor do they correlate well with other evaluative tools that do measure such ability.[6,11,12,19]

The OSCE is a relatively new tool for evaluating physicians in training. The examination was introduced by Harden et al. at the University of Dundee (Dundee, Scotland) in an effort to improve the evaluation of medical students' clinical performance.[7] As described by the authors, the OSCE consisted of a number of stations through which medical students rotated, spending predetermined amounts of time at each station. Different clinical skills were assessed at the individual stations, and item checklists were used to objectively grade students' skills in performing clinical tasks. As elaborated by Harden et al., the OSCE offered the advantages of controlled grading criteria and easy repeatability of the examination.[7] Subsequent authors have shown the OSCE to be a reliable and valid examination not only for medical students but also for residents in a variety of disciplines, including internal medicine and surgery.[8–12] In Canada, confidence in the OSCE as a superior method of performance evaluation has led the Medical Council of Canada and several specialty boards to include the OSCE in the licensing and certifying examination process.[13–15] Although such steps have not been taken in the United States, there is considerable interest in expanding the use of performance-based testing to evaluate the performance of medical students and residents. We believe that it is crucial that the OSCE be completely evaluated for reliability and validity before the numerous resources (*e.g.*, money, faculty time, patient recruitment and training, administrative costs, quality control costs, etc.) necessary to develop OSCE examinations for certification and licensure are expended. One of the primary problems with the OSCE is the wide range of reliabilities, from 0.19 to 0.89, reported for this examination.[9,10,14–18] This wide range has actually led some authors to suggest that the test is only reliable if it lasts for 6 to 10 hours, which would clearly produce an unmanageable and impractical examination.[17] It is absolutely necessary to show that the OSCE has a reliability that can justify its use as a licensing examination. In addition, to our knowledge, an OSCE has never before been given to all residents in an entire residency program. Administering the OSCE in this manner is mandatory if we are to determine the validity of such an examination. It is crucial to demonstrate that the OSCE can detect significant differences in clinical performance between residents at different levels of training.

The reliability of this OSCE examination was excellent (0.91), exceeding both the accepted benchmark reliability standard of 0.80 and all other reported reliabilities. This level of reliability is comparable to the reliability of the multiple-choice examinations that are used for the purposes of the In-Training Examination and specialty board certification.[25,26] A testing time of 3 hours was

sufficient to achieve reliability. In fact, the Spearman–Brown equation indicated that an even shorter examination containing only 16 stations or 8 problems would have been associated with an acceptable level of reliability. We find it interesting that in the literature, reliabilities for the OSCE appear to be better for residents than for medical students. We attribute the high reliability of the OSCE in the current study partly to the population tested, which consisted of residents at multiple levels of training. The reliability was also improved by the large number of clinical problems.

Another advantage of testing residents at multiple training levels is that validity can be established. Previous reports on the reliability of the OSCE have not been based on the examination of an entire group of residents. A valid examination should show that more experienced residents perform at a higher level. In the current study, the performance of residents could be grouped into three categories, the highest level of performance coming from the residents at the most senior levels. Joorabchi used an OSCE to evaluate a group of pediatric residents and found it to be more effective at identifying the level of residency training than either the Pediatric Board In-Training Examination or resident performance ratings.[12] Cohen et al., at the University of Toronto, used the OSCE to evaluate the clinical skills of PGY2 surgical residents.[9] Significant differences in scores were found between these residents and foreign medical graduates.[9]

The question of whether the OSCE provides information not provided by other testing methods, such as the ABSITE, cannot be answered, because scores on both testing methods tend to increase with level of experience. Thus, the two methods will tend to rank individuals in the same way even though they may be measuring very different skills. The simple correlations between ABSITE clinical scores and OSCE scores demonstrated just such an effect. However, when level of training was controlled for, there was a dramatic decrease in the magnitude of the correlations between ABSITE and OSCE. Nevertheless, the ABSITE clinical scores did correlate significantly with the OSCE scores, and to some limited degree they measure similar skills.

The OSCE has not been used as a screening test for surgical residents by our institution or by any other, to our knowledge. The weak performance of incoming interns in general is disturbing and underscores significant deficiencies in undergraduate medical education. With the growing number of institutions using the OSCE method to test medical students, perhaps program directors should look specifically at the OSCE scores in surgery and in other disciplines of prospective residents.

The advantages of the OSCE are numerous. By definition, residents are placed in well-defined clinical scenarios in which the variables can be controlled. There is no limit to the variety of clinical situations that can be constructed. Although most authors have relied almost exclusively on persons playing the role of patients (i.e., "simulated" patients), we have attempted to use as many actual patients as possible. Although it has been shown that simulated patients reliably mimic actual patients,[27] actual patients are clearly superior in demonstrating physical findings, such as a thyroid nodule, a venous stasis ulcer, or an ischemic foot.

The OSCE allows residents to be placed in clinical situations that cannot be duplicated by a case presented on paper. Further, the OSCE format permits evaluation of more obscure areas, such as the residents' abilities to interact effectively with a patient.[28] We have previously demonstrated a correlation between resident interpersonal skills and clinical performance on the OSCE.[28] Without question, a resident with poor interpersonal skills stands out in the OSCE, and we have shown that faculty members can reliably measure this subjective aspect of a resident's performance with this test.[28]

A real advantage of the OSCE is that it demonstrates to residents that basic clinical skills are valued by faculty members. Having a faculty member directly observe residents either taking a history or performing a physical examination underlines the importance of clinical examination and sends a clear message to the residents. Not only are faculty rating forms vague, but also, faculty members are reluctant to criticize trainees on the basis of this subjective information.[29] The OSCE results are more objective and allow for much more accurate feedback to residents. Such feedback is particularly valuable when dealing with the problem resident. Typically, the only information available for judging a resident who is performing poorly is gained from faculty ratings, which, as previously stated, tend to underestimate the problem.[3,19,29] The OSCE provides the program director with much more objective information than would otherwise be available.

Perhaps the greatest benefit of the OSCE is that it allows identification of problem areas.[11,20,21] The OSCE can be used to identify weaknesses within the resident curriculum. In the current study we showed numerous deficits in clinical performance on basic clinical problems, even among the more senior residents. We have begun to use this information to modify our curriculum, as have other investigators.[30]

We conclude that the OSCE is a very reliable and valid method for evaluating residents. Because the OSCE provides a unique insight into the progression of residents' clinical competence, we believe that it should become a standard part of resident evaluation. We believe that the data support our contention that the OSCE is the new gold standard for evaluating the clinical performance of residents.

References

1. Maxim BR, Dielman TE. Dimensionality, internal consistency, and interrater reliability of clinical performance ratings. Med Educ 1987; 21:130–137.
2. Ansell JS, Boughton R, Cullen T, et al. Lack of agreement between subjective ratings of instructors and objective testing of knowledge acquisition in a urological continuing medical education course. J Urol 1979; 122:721–723.
3. Schwartz R, Donnelly M, Drake D, Sloan D. Faculty sensitivity in detecting medical students' clinical competence. Clin Invest Med 1993; 16(suppl):B87.
4. Goetzl EJ, Cohen P, Downing E, et al. Quality of diagnostic examinations in a university hospital outpatient clinic. JAMA 1983; 249:1035–1037.
5. Wray NP, Friedland JA. Detection and correction of house staff error in physical diagnosis. Ann Intern Med 1973; 78:481–489.
6. Levine HG, McGuire CH. The validity of multiple choice tests as measures of competence in medicine. Am Educ Res J 1970; 7:69–83.
7. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J 1975; 1:447–451.
8. Petrusa ER, Blackwell TA, Carline J, et al. A multi-institutional trial of an objective structured clinical examination. Teach Learn Med 1991; 3:86–94.
9. Cohen R, Reznick RK, Taylor BR, et al. Reliability and validity of the objective structured clinical examination in assessing surgical residents. Am J Surg 1990; 160:302–305.
10. Petrusa ER, Blackwell TA, Ainsworth MA. Reliability and validity of an objective structured clinical examination for assessing the clinical performance of residents. Arch Intern Med 1990; 150:573–577.
11. Sloan DA, Donnelly MB, Johnson SB, et al. Use of an objective structured clinical examination (OSCE) to measure improvement in clinical competence during the surgical internship. Surgery 1993; 114:343–351.
12. Joorabchi B. Objective structured clinical examination in a pediatric residency program. Am J Dis Child 1991; 145:757–762.
13. Reznick RK, Blackmore D, Cohen R, et al. An objective structured clinical examination for the licentiate of the medical council of Canada: from research to reality. Acad Med 1993; 68(10):S4–S6.
14. Reznick R, Smee S, Rothman A, et al. An objective structured clinical examination for the licentiate: report of the Pilot Project of the Medical Council of Canada. Acad Med 1992; 67:487–494.
15. Grand'Maison P, Lescop J, Rainsberry P, Brailovsky CA. Large scale use of an objective structured clinical examination for licensing family physicians. Can Med Assoc J 1992; 146:1735–1740.
16. Matsell DG, Wolfish NM, Hsu E. Reliability and validity of the objective structured clinical examination in paediatrics. Med Educ 1991; 25:293–299.
17. Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. Med Educ 1988; 22:325–334.
18. Roberts J, Norman G. Reliability and learning from the objective structured clinical examination. Med Educ 1990; 24:219–223.
19. Schwartz R, Donnelly M, Sloan D, et al. The relationship between faculty ward evaluations, OSCE, and ABSITE as measures of surgical intern performance. Am J Surg (in press).
20. Sloan DA, Donnelly MB, Schwartz RW, et al. Assessing the clinical competence of medical students and surgery residents in surgical oncology problem solving. Ann Surg Oncol 1994; 1:204–212.
21. Endean E, Sloan D, Veldenz H, et al. Performance of the vascular physical examination by residents and medical students. J Vasc Surg 1994; 19:149–156.
22. Gleeson F. Defects in postgraduate clinical skills as revealed by the objective structured long examination record (OSLER). Ir Med J 1992; 85:11–14.
23. Reisner E, Dunnington G, Beard J, et al. A model for the assessment of students' physician–patient interaction skills on the surgical clerkship. Am J Surg 1991; 162:271–273.
24. Stillman PL, Swanson DB, Smee S, et al. Assessing clinical skills of residents with standardized patients. Ann Intern Med 1986; 105: 762–771.
25. Grosse ME, Craft GE, Blaisdell W. The American Board of Surgery In-Training Examination. Arch Surg 1980; 115:654–657.
26. Langdon LO, Grosso LJ, Day SC, et al. A core component of the certification examination in internal medicine. J Gen Int Med 1993; 8:497–501.
27. Norman GR, Tugwell P, Feightner JW. A comparison of resident performance on real and simulated patients. J Med Educ 1982; 57: 708–715.
28. Sloan D, Donnelly M, Johnson S, et al. Assessing the interpersonal skills of surgical residents and medical students. J Surg Res 1994; 57:613–618.
29. Irby DM, Fantel JI, Milam SD, Schwarz MR. Special report: legal guidelines for evaluating and dismissing medical students. N Engl J Med 1981; 3:180.
30. Lunenfeld E, Weinreb B, Lavi Y, et al. Assessment of emergency medicine: a comparison of an experimental objective structured clinical examination with a practical examination. Med Educ 1991; 25:38–44.